Journal of Forestry Research (2011) 22(2): 201–204
DOI 10.1007/s11676-011-0150-3

ORIGINAL PAPER

Identification of SSR loci in *Betula luminifera* using birch EST data

LU Yong-quan • LI Hai-ying • JIA Qing • HUANG Hua-hong • TONG Zai-kang

Received: 2010-08-03; Accepted: 2010-09-30

Abstract: Expressed sequence tags (ESTs) are generated from single-pass sequencing of randomly picked cDNA clones and can be used for development of simple sequence repeat (SSR) markers or microsatellites. However, EST databases have been developed for only a small number of species. This paper provides a case study of the utility of freely available birch EST resources for the development of markers necessary for the genetic analysis of *Betula luminifera*. Based on birch EST data, primers for 80 EST-SSR candidate loci were developed and tested in birch. Of these, 59 EST-SSR loci yielded single, stable and clear PCR products. We then tested the utility of those 59 markers in *B. luminifera*. The results showed 28 (47.6%) yielded stable and clear PCR products for at least one *B. luminifera* genotype. In addition, this study describes a rapid and inexpensive alternative for the development of SSRs in species with scarce available sequence data.

Keywords: exploit; *Betula luminifera*; birch; EST database; EST-SSR

Introduction

Molecular markers have broad uses in genetic research (e.g. construction of genetic maps and gene mapping), breeding (e.g. molecular marker-assisted selection), gene cloning, and comparative genomics. Many types of molecular markers have been developed since restriction fragment length polymorphisms (RFLP) were developed in 1980 (Botstein et al. 1980). Up to now the more powerful and available markers are those based on PCR techniques. In short, there are two types of PCR-based markers; one is random primer markers which can be used in nearly all kinds of species, examples are randomly amplified polymorphic DNAs (RAPDs) (Williams et al. 1990) and amplified fragment length polymorphisms (AFLPs) (Vos et al. 1995). The other is special primer markers such as SSR (simple sequence repeat), which must be developed from and used in target species (Becker et al. 1995). Random primer markers are more flexible because they can be used in nearly all kinds of species. But their reliability, especially in RAPDs, is in doubt in some extent.

SSRs have proven to be more reliable than other markers. SSRs consist of tandem repeats of short (1–6 bp) nucleotide motifs (Gupta et al. 1996). These repeat sequences are distributed throughout the genome. Polymorphism revealed by SSRs results from variation in repeat number, which primarily results from slipped-strand mispairing during DNA replication. Thus, SSRs reveal much higher levels of polymorphism than most other marker systems (Toth et al. 2000; Li et al. 2002). Although the utility of SSRs in genetics studies is well established, the isolation and characterization of such markers by traditional methods is costly and time consuming, which makes the *de novo* development of SSRs unrealistic for some taxa (Becker et al. 1995; Pashley et al. 2006).

Expressed sequence tags (ESTs) are generated from single-pass sequencing of randomly picked cDNA clones (Adams et al. 1991). The EST approach and subsequent gene-expression profiling (cDNA microarrays) have proven to efficiently identify genes and analyze their expression during different developmental stages or under various environmental stresses (Fowler et al. 2002; Wei et al. 2005). With the recent progress made in large-scale plant function genome sequencing project, thousands of data sets have been generated, and importantly, most of these are freely available for use by any plant biologist word wide (Brady et al. 2009). These ESTs are useful for developing SSR markers (EST-SSR). Thus, the use of such database to develop SSRs could be an inexpensive and rapid alternative to traditional methods (Wang et al. 2005). At present there remain many species with no EST data. At the same time, those species need more efficient molecular markers to support research into their biology.

Foundation project: This work was funded by the Provincial Natural Science Foundation of Zhejiang in China (Y307465).

The online version is available at http://www.springerlink.com

LU Yong-quan (✉) • LI Hai-ying • JIA Qing • HUANG Hua-hong TONG Zai-kang
The Nurturing Station for the State Key Laboratory of Subtropical Silviculture Zhejiang Agriculture and Forestry University Lin'an 311300, China. Email: luyongquan@126.com

Responsible editor: Hu Yanbo

According to Yang et al. (2007), EST-derived markers are likely to be conserved across a broader taxonomic than any other sorts of marker. For those species with scarce sequence information, it may be feasible to develop EST-SSR markers using EST data for closely related taxa (Lu et al. 2006). This study explores that strategy.

Pashley et al. (2006) described a case study in which the publicly available cultivated sunflower EST database was used to develop SSR markers for use in the genetic analysis of another two sunflower species. The result showed that EST-derived SSR were more than three times as transferable across species as compared with genome SSR. Moreover, EST-SSRs whose primers were located within protein-coding sequence were more readily transferable than those derived from untranslated regions. This survey revealed that more than one-third of all plant-derived EST collections of sufficient size could conceivably serve as a source of EST-SSRs for the analysis of rare, endangered, or invasive plant species worldwide.

Up to now, tree biology research, in contrast with the others, lags behind. The databases of tree sequence data are limited. For many tree species, there is no DNA sequence information. This lack of data makes it difficult to develop locus specific primers for those species. There is narrow choice of markers for biology research in those species. It is useful and necessary to develop and test more markers for those species.

Betula luminifera is a deciduous tree, widely occurring in temperate zone in China. As lots of other tree species, its biology and sequence data are limited. Birch (*Betula platyphylla* Suk), which has lots of sequencing and biological information in public databases, is closely related to *B. luminifera*. This paper provides a case study of the utility of freely available birch EST resources for the development of markers necessary for genetic analyses of *B. luminifera*.

Materials and methods

Search of putative SSR

In total, 3 028 (*B. platyphylla*) EST sequences of birch, released by the plant GDB (http://www.plantgdb.org) were examined. We used SSR Finder (http://www.gramene.org.gramene/searches/ssrtool) to search SSR among these ESTs. Those ESTs including a 2–4 bp repeat motif were select as putative SSR.

Targeting of candidate SSR by electronic PCR

To develop candidate SSR markers from the SSRs identified with SSR Finder, we designed PCR primers based on flanking regions on the EST sequences using ePrimer3 (http://www.hgmp.mrc.ac.uk). For convenience, we used a 200-bp cDNA sequence with 100 bp on each side of the target split for the primer design for each EST-SSR. Then we tested the designed primers by electronic PCR (e-PCR; Schuler 1997) on the birch EST sequence. To increase the quality and usability of the in silico exploited SSR markers, we required exact matches between primers and templates and set a 600-bp length on the product size for the e-PCR. We accepted a putative EST-SSR locus as a candidate EST-SSR marker only if those primers successfully and uniquely amplified the correct target in the e-PCR. The candidate primers were selected and named with the abbreviation BES (for *B. luminifera* EST-SSR) followed by a unique number (e.g. BES 18).

Verification and evaluation of SSR markers in *Betula luminifera*

One birch, as well as two *B. luminifera* varieties including Lin'an5 and Sichuan4 which represent two different ecotypes, were used to verify the candidate SSR markers. Total genomic DNA was isolated from 200 mg of fresh leaf tissue using CTAB method (Murray et al. 1980).

All primers used were synthesized by Nanjing Jinsite Biological Engineering & Technology Company in China. PCR was performed in 20 μL reactions containing 50 ng of template DNA, 0.5 μmol/L of each primer, 200 μmol/L of each dNTP, 1.5 mmol/L of $MgCl_2$, 1 unit of Taq polymerase, and 2 μL of 10 × PCR reaction buffer. A touchdown PCR program (Don et al. 1991) was used: 5 min at 95°C; 10 cycles of 30 s at 95°C, 30 s at 58°C minus 0.3°C/cycle, 1 min at 72°C; 20 cycles of 30 s at 95°C, 30 s at 55°C, 1 min at 72°C; and 7 min at 72°C for a final extension. For those primer pairs that did not generate good amplification results, the initial annealing temperatures were adjusted from 55°C to 60°C. Each of the primer pairs was tested twice to confirm the repeatability of the observed bands in each genotype. PCR products were separated on agarose gel. Gels were stained with Ethidium Bromide for visualizing DNA bands.

Sequencing PCR product

To confirm the PCR product amplified in *B. luminifera* were homologous to the birch genes where the loci were first identified, a band yielded by primer BES17 in Lin'an was isolated, purified and sequenced.

Results and discussion

Candidate SSR markers

By screening 3 028 EST sequences from birch with SSR Finder, we identified 331 ESTs carrying SSR motifs. These results showed that about 10.9% ESTs have SSR, we successfully obtained 151 (45.6%) e-PCR products from those 331 putative SSR loci. Those that did not yield a product were the result of too short sequences flanking the SSRs. With these 151 primers, we successfully obtained 333 e-PCR products. This result showed that there are multiple BAC clones or overlap of ESTs obtained same PCR product. Multiple-copy markers are not desirable for genetic studies so we discarded these primer pairs. At last, 80 candidate SSR markers were selected for further development.

Experimental tests of candidate SSR markers

Because the primers were designed from birch EST data, they must be tested for application in *B. luminifera* by experiment. All primers were tested on birch DNA preparations first (Fig. 1). Of the 80 candidate SSR markers tested, 68 (85%) yielded stable and clear PCR products as expected in birch. We suspect that the main reason for amplification failure of the remaining 12 candidate loci is that the species of birch used in this study is different from that from which the EST data were generated. Among the 68 candidate SSR markers tested, 59 (86.8%) yielded single, stable and clear PCR products as expected in birch. We then tested the utility of those 59 markers in *B. luminifera* (Fig. 2). The results showed 28 (47.6%) yielded stable and clear PCR products in at least one *B. luminifera* genotype. The primers for these new *B. luminifera* SSR markers and their names are shown in Table 1.

Table1. Twenty-eight EST-SSR markers developed for *Betula luminifera*

Primer name	ID in PlantGDB[16]	Former primer 5-3	Reverse primer 5-3
BES2	gi|34389440	GCCGGGGAAGAAAGTTACC	CACGTTGGGAATGTGATGAT
BES3	gi|34389519	CCAACAGGCTTTCATTTGCT	ATCAGGGGCATCAACAAGAG
BES7	gi|34388540	TCTCTTCCCGCAAACTCTCT	ATAAACCGCCCAGGAAAAAC
BES10	gi|34389631	GAATGTTCTCTGCTCCTCCAG	TCACTATTCGGTGCAACAGG
BES15	gi|34389670	CCCCCTCCCTTTTACTCTTTC	TTCTGCTCCCGTCTCATCTT
BES17	gi|34388669	TCTCACCAAACCACTCACTCA	AAGAGCGTGGCAATGAACTC
BES18	gi|34389738	CAGACGACAAAGCAAGCTGA	CATGCTCACATACAAGGCAAA
BES19	gi|34389809	GCGACACACCCTACCATCTT	GGTGCACTTGCAGATGTGAT
BES20	gi|34388674	GGTTGCTCAACCTAACCAACA	AGAACACCCACCAAGTCACC
BES22	gi|34389885	AGGGTGTTCAAACCGACGA	CGGTCTCAATCTCCACGTTT
BES23	gi|34389943	GCACTCACTCGGATACTCGTC	CTTTTGCACCATGTTTGTGG
BES24	gi|34389962	GCCGGGAGAATTACACGTC	CCCCTTTCTTCAGATCAACG
BES25	gi|34388873	TAGAGCGTTGCGCAGATAGA	CAGGTTCCTCTCCTCCACTG
BES27	gi|34390086	TTTATTTTTCAATTTTTCCTAGAGAGG	ACCACACCGAGGCATACAAT
BES29	gi|34390130	GCGACAGGAAATTCAACCAC	CTGCGTCAGACTGCACATTT
BES31	gi|34388686	TGAATAGACCGTTGCGCTTA	CGTATCTCTCGGCTTGCTCT
BES33	gi|34390195	GAGAGAACCAAAACAGTAGACAGAGA	GGCCTGTTCTTGATGACGAT
BES35	gi|34390304	GGGGGTTGCTCTTCATTTTT	GGTTTCCTCGTCGGTTATGA
BES36	gi|34388710	CGCCAAATCTTTACCCAGAA	CGACGATGATGATCCATGAG
BES39	gi|34390489	CGGGGGACATTACAAATAGC	TCGCATCTTCATCTGTGAGG
BES40	gi|34390566	CCCTGCCTCTTCTCTGTCAC	GCCATAAGCCTCCAATCTCA
BES45	gi|34390675	CAGCGTATGAAACCAGAACG	TAAAACGGACCCACTTGAGC
BES48	gi|34390740	GTTAAGAAGGTGCGCCAGTC	ACTAACCGCGCATAAACTGC
BES60	gi|34388842	GGACTTCTTCGGAGACATGG	CCCCAGAAAATAACGGCATA
BES62	gi|34388871	CCTCTCTGGCTTCTCCTCCT	TCGAATCCATATCCACCAAAA
BES67	gi|34389051	CGGGGGAACCATCAAAAA	CGGGAAGTCGCATATAGGAA
BES70	gi|34389144	GGCGTTTAATCTGGGTGAGA	ACGCCAGAATGGTAGACACC
BES80	gi|34389429	TCAGCTTCGTTCCAAAACCT	CCCATTTGGAGATGGAGAAA

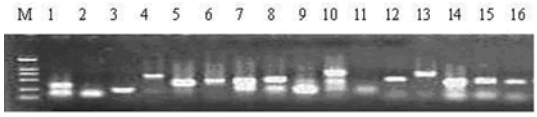

Fig. 1 Some of the primers screened in birch separated by electrophoresis on 1% agrose gel (M Marker DL2000)

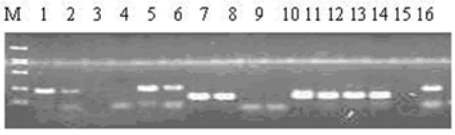

Fig. 2 Some of the primers screened in *B. luminifera* separated by electrophoresis on 1% agrose gel (M DNA Marker DL2000). Each two neighboring lanes (from 1 to 20) are separated by the same prime. Odd lanes are PCR products from Lin'an5 DNA templet; even lanes are PCR products from Xichuan4 DNA templet

Homologous test

To examine the homology of SSR markers amplified in *B. luminifera* using primers derived from the EST sequence in birch, we randomly selected one PCR product yielded by primer BES17 in the Lin'an variety for DNA sequencing. BLAST results of the 177 bp fragment (Fig. 3) showed that the sequence shared 88% identity with the EST (gi|34388669) that generated the SSR primer (BES17), which met our expectations (Fig. 4).

CCCTTCCTCTAGTTAGCGCGGGTGCCATGGCTAAAGACATGGAAAGTTGGAGGGCAACGT
GGTTTCTCCCCCAAGGACTACCACAACCCACCACCGGCGCCGCTGATTGGTGCACACGAG
CTCACAATGTGGTCCTTCTACCGGTCTATCATTGCCTACCTCATTGCTACGCTCATA

Fig. 3 Nucleotide sequence of PCR produce yielded by primer BES17

Fig. 4 Nucleotide alignment between PCR product of BES17 and EST (gi|34388669)

Good molecular markers are useful tools for genetic and biology studies. The study of plant biology, as with all areas of biology, has undergone dramatic changes in the past decade since the development high-throughput methods for sequence determination. In recent years, high-density oligonucleotide re-sequencing microarrarys and next-generation sequencing technologies have resulted in a considerable increase in the amount of available genome sequence data (http://www.ncbi.nlm.nih.gov). However, there still lots of species, especially woody plants, have little for no publicly available sequence information. And at present, it is not feasible to sequence the genome of all the species. However, we verified in this paper, EST data could work very well for developing genetic tools for taxa closely related to EST sequencing targets. We verified our primers in a large population within the *Betula* and they worked well (detailed results are the subjects of another article). With the help of these results, it is possible to follow similar strategies to develop SSR primers in a variety of species with scarce sequence data. Making EST-SSR marker could be a good choice for those species.

In addition, because EST-SSRs are genetic markers residing in gene sequences, they can directly reflect aspects of variation within those genes. Therefore, the maps constructed with EST-SSR markers could be specially valuable for genetic studies. On the other hand, levels of polymorphism could be low because these SSRs are in expressed regions which may have more evolutionary constrains compared with SSRs derived from untranscribed sequences. Nevertheless, EST-SSR marker can be useful for comparative genomic studies precisely because they are designed in expressed regions. Therefore it could be used in wide range, especially the analysis of linear relationship among different genomes within same genus (Rong et al. 2004; Lu et al. 2006).

References

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC. 1991. Complementary DNA sequencing: expressed sequence tag and human genome project. *Science*, **252**: 1651−1656.

Becker J, Heun M. 1995. Barley microsatellites: allele variation and mapping. *Plant Mol Biol*, **27**: 835−845.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*, **32**: 314−331.

Brady SM, Provart NJ. 2009. Web-queryable large-scale data sets for hyposis generation in plant biology. American society of plant Biologist, in web http://www.aspb.org.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res*, **19**: 4008.

Fowler S, Thomashow MF. 2002. Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell*, **14**: 1675−1690.

Gupta PK, Balyan IS, Sharma PC, Ramesh B. 1996. Microsatellites in plants: a new class of molecular markers. *Curr Sci*, **70**: 45−54.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol*, **11**: 2453−2465.

Lu Yongquan, Wang Xusheng, Huang Weisu, Xiao Tianxia, Zheng Yan, Wu Weiren. 2006a. Development of amplified consensus genetic markers in Gramineae based on rice intron length polymorphisms. *Scientia Agricultura Sinica*, **39**: 433−439. (in Chinese with an English abstract).

Lu Y, Ye Z, Wu W. 2006b. Analysis of the phylogenetic relationships among several species of gramineae using ACGM markers. *Acta Genetica Sinica*, **33**: 1127−1131.

Murray MG, Thompson WF. 1980. Rapid isolation of high molecular-weight plant DNA. *Nucleic Acids Res*, **8**: 4321−4325.

Pashley CH, Ellis JR, Mccauley DE, Burke JM. 2006. EST Database as a source for molecular markers: lessons from Helianthus. *Journal of heredity*, **97**: 381−388.

Rong J, Bowers JE, Schulze SR, Waghmare VN, Rogers CJ, Pierce GJ, Zhang H, Estill JC, Paterson AH. 2005. Comparative genomics of Gossypium and Arabidopsis: unraveling the consequences of both ancient and recent polypolidy. *Genome research*, **15**: 1198−1210.

Schuler GD. 1997. Sequence mapping by electronic PCR. *Genome Res*, **7**: 541−550.

Toth G, Gaspari Z, Jurka J. 2000. Microsatellites in different eukaryotic genomes: survey an analysis. *Genome Res*, **10**: 967−981.

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res*, **23**: 4407−4414.

Wang X, Zhao X, Zhu J, Wu W. 2005. Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.). *DNA Research,* **12**: 417−427.

Wei H, Dhanaraj AL, Rowland LJ, FU Y, Krebs SL, Arora R. 2005. Comparative analysis of expressed sequence tags from cold acclimated and non-acclimated leaves of *Rhododendron catawbiense* Michx. *Planta*, **221**: 406−416.

Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. *Nucleic Acid Res*, **18**: 6531−6535.

Yang L, Jin G, Zhao X, Zheng Y, Xu Z, Wu W. 2007. PIP: a database of potential intron polymorphism markers. *Bioinformatics*, **23**: 2174−2177.